보건학석사 학위논문

A Genome-Wide Association Study on Hallux Valgus: The Healthy Twin Study Korea

무지외반증과 관련된 유전요인 및 관련요인 분석

2014년 08월

서울대학교 보건대학원

보건학과 유전체역학전공

이 수 지

보건학석사 학위논문

A Genome-Wide Association Study on Hallux Valgus: The Healthy Twin Study Korea

무지외반증과 관련된 유전요인 및 관련요인 분석

2014년 08월

서울대학교 보건대학원

보건학과 유전체역학전공

이 수 지

Abstract

A Genome-Wide Association Study on Hallux Valgus: The Healthy Twin Study Korea

서울대학교 보건대학원

보건학과 유전체역학전공

이 수 지

Introduction: Hallux valgus (HV) is one of the most common chronic structural foot deformities to be associated with genetic predisposition, age, sex, and footwear. Today, growing tendency to reach the ideal of female perfect appearance explains the tacit social pressure on wearing high heels, which may aggravate HV. Several studies have found that there is an association between constricting footwear and HV. Given the familial patterns of foot structure, several studies have been conducted to indicate genetic factors play a role. However, specific genetic variants underlying HV have not yet been identified. The objective of this study is to investigate common genetic variants that confer susceptibility to the hallux valgus deviation in HV deformities through a genome-wide association (GWA) study. **Methods:** The Healthy

Twin study is an ongoing cohort of adult same-sex twin pairs, aged ≥30 years, and their first-degree family members who have been recruited through participating hospitals since 2005. Among the 3461 total, 1265 individuals have taken the weight-bearing anteroposterior (AP) and lateral foot X-ray examinations. Out of those, 995 individuals (628 women) with genetic information were finally included for the GWA study. Hallux valgus angle (HVA) was obtained from each foot X-ray, and information about narrow-toed and/or high-heeled shoe use was collected through self-reported surveys. The software Merlin, available approach for monozygotic (MZ) twins included family-based data, was performed to interrogate association between the trait of interest and genetic variants. **Results:** potential candidate single nucleotide polymorphisms(SNP) were identified. And, the regions including those SNPs comprise genes of DLX5, NTNG1, PRMT6, XYLT1, and CCDC85A, which may be a potential target for hallux valgus treatment and general foot development. **Discussion:** this study successfully identified potential genetic variants influencing HVA deviation. These findings open up for further validation through replication studies, and are expected to guide future studies for foot health.

Keywords: Hallux Valgus, Genome-wide association study, Single nucleotide polymorphism, Genetic variants, Association testing with related individuals

Student Number: 2012-21895

Contents

List of Tables

List of Figures

I. Introduction

A. Background

Historically, female foot health has been an issue since the 10th century when foot-binding was a prevailed tradition - for females of Han nationality, till now. Chinese girls had their feet wrapped by their mothers at about the age of 6 to prevent the feet from maturing normally and to create as small feet as possible so as to greatly enhance the girls' matrimonial prospects. Bound feet were the symbol of the wealth and girls with small feet were preferred by the Chinese men. However, due to foot-binding, a girl's feet would turn atrophic, resulting in deformation, pain, and mobility restriction [1-2].

Drawn into modern times, Hallux Valgus (HV), one of the most common chronic structural foot deformities, characterized by the progressive abnormal angulation of the first metatarsophalangeal joint due to lateral deviation of the big toe and medial deviation of the first metatarsal head (Figure 1), is affecting 20.6% of the Korean female population; and its prevalence in Koreans' general population shows 16.4%[3]. Additionally, approximately 23% in older adults across different ethnic populations are affected[4], which pins down that the deformity grows as a menace to public health.

Previous studies have implicated age, female sex, higher body mass index (BMI), constricting footwear[5-7], and genetic predisposition{[3,8] are associated with HV: genetics has been

suspected as a risk factor in clinical settings where there is a tendency for those with structural foot deformities to show a family history[9-13]. These studies estimated 63-90% of HV patients have other affected family members.

Today, growing social tendency to reach the ideal of female perfect appearance explains the tacit social pressure on wearing high heels, aggravating HV and leading to HV pain and difficulty in walking, of which such patients require orthosis as a result. Given the familial patterns of foot structure, several studies have been conducted to indicate genetic factors play a role, indicating the heritability estimates[3,8]. However, no study has yet identified specific genetic variants underlying HV. Verifying which variations are associated with the condition of interest would assist in identifying the high-risk group for early intervention.

Figure 1. Hallux Valgus

B. Previous studies

Genetic evidence for hallux valgus has been previously studied with high familiar correlation and heritability. The intra-class correlation coefficient (ICC) a statistic that describes how strong resemblance members in the same group show. In the previous study, ICC was to quantify the extent to which individuals with a fixed degree of relatedness - monozygotic twin pairs, parent-offspring pairs, dizygotic twin and full sibling pairs - resemble each other. The higher the ICC, the more resemblance in the trait. Table 1 describes the estimate of each relation group.

Table 1. Intra-class Correlation Coefficient estimates of Hallux Valgus Angle and Status for each group.

	Intra-class Correlation Coefficients (95% CI)		
	MZ twin pairs (n=175)	parent-Offspring pairs (n=959)	DZ and sibling pairs (n=574)
HVA	0.53 (0.41-0.63)	0.10 (0.03-0.16)	0.27(0.20-0.35)
HV (HVA>20°)	0.51 (0.41-0.63)	0.16 (0.10-0.23)	0.37(0.30-0.44)

*Note: CI, confidence interval; MZ, monozygotic; DZ, dizygotic

Heritability is measured as the phenotypic variance attributed to genetic variance, where the phenotypic variance is due to genetic and/or environmental variability, as below.

$$\text{Phenotype (P)} = \text{Genotype (G)} + \text{Environment (E)}$$

The function for phenotypic variance,

$$\text{Var(P)} = \text{Var(G)} + \text{Var(E)} + 2\text{Cov(G,E)},$$

can be converted as below, with genetic-environmental covariance held as zero.

$$\text{Var(P)} = \text{Var(G)} + \text{Var(E)}$$

The broad-sense heritability, H^2 is defined as relative contribution of all sorts of genetic variance to the total phenotypic variance.

$$H^2 = \text{Var(G)}/\text{Var(P)}$$

Among the possible genetic variances, that are additive, dominant, epistatic, and parental effects, in particular, the additive variance is important, so that narrow-sense heritability, h^2 can be now defined as

$$h^2 = \text{Var(A)}/\text{Var(P)}$$

Furthermore, the environmental effect also can be considered in the model, which can be decomposed into common household (C) and random non-familial (R) environmental variances.

$$Var(P) = Var(G) + Var(C) + Var(R)$$

Each different type of model underlying the heritability estimation can lead different results. In the previous study, two models were tested - AE and ACE models. The model fitness was compared according to Akaike Information Criterion (AIC) to select a more fitted one.

AIC = 2k - 2ln(L), where k is the number of parameters used in the model, and L is the maximized estimate of the likelihood function for the model. Less AIC score indicates better fitness. Each model tested is given below in Table 2.

Table 2. Heritability Estimates of Hallux Valgus Angle and Status

	Variance composite model (95% CI)				
	A	C	E	Model (AIC)	Variation explained by covariates*
HVA	0.47(0.38-0.56)	0	0.53(0.44-0.62)	ACE (5821.8)	0.06
	0.47(0.38-0.56) p=4.52e-28		**0.53(0.44-0.62)**	**AE (5819.8)**	
HV (HVA>20)	0.51(0.42-0.59)	0	0.49(0.30-0.68)	ACE (1043.4)	0.06
	0.51(0.42-0.59) p=5.80e-09		**0.49(0.30-0.68)**	**AE (1041.4)**	

*Covariates include age, sex, age*sex, age^2, age^2*sex.

C. Purpose

The objective of this study is to investigate genetic variants that confer susceptibility to the deviation in hallux valgus angle through a genome-wide association study (GWAS).

II. Methods

A. Study population

The Healthy Twin Study is an ongoing cohort of same-sex adult twin pairs, aged ≥30 years, and their first-degree family members who have been recruited through participating hospitals since 2005 and followed up every 3 years. Ascertainment, recruitment, protocols and general characteristics of participants are described in detail previously[14]. Among the 3461 participants total, 1265 individuals who have taken the weight-bearing anteroposterior (AP) and lateral foot X-ray examinations. Finally, 996 out of the 1265, with genetic information were included in genome-wide association analysis. These consisted of 234 families, and 152 monozygotic (MZ) twin pairs.

B. Measurement

Hallux valgus angle (HVA) was obtained from each foot radiograph taken in the AP standing position on a cassette, and information about participants' current footwear habits for narrow-toed and/or high-heeled shoes were collected through self-reported surveys. Presence of hallux valgus as a dichotomous trait was evaluated based on a trained physiotherapist's visual inspection on the X-ray screening

results. Though, considered to be present if the angle of the big toe towards the other toes was observed to be larger than 20°, this study focused on the angular deviation extent in degrees as a quantitative trait (Figure 2). Known risk factors such as age, sex, and body mass index (BMI) were used as covariates in analyses. For BMI estimation, weight(kg) and height(cm) were measured in light clothing using standardized scales and stadiometer; then, computed dividing the weight by the squared height.

Figure 2. Measurement of Hallux Valgus Angle (HVA)

C. Genotype data

Genomic DNA was extracted from whole-blood samples obtained from subjects at their recruitment and genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, CA, USA). Genotype calling was completed using the Birdseed v2 algorithm, which performs a multiple-chip analysis to estimate signal intensity for each allele of each SNP, fitting probe-specific effects to increase precision. Then, it makes genotype calls by fitting a Gaussian mixture model in the two-dimensional A-signal vs. B-signal space, using a SNP-specifically customized Expectation-Maximization (EM) clustering algorithm to improve accuracy measured in confidence scores for every individual at every SNP.

Exclusion criteria for single nucleotide polymorphism (SNP) performance included:

(1) duplicate SNPs
(2) call rate < 95%
(3) Hardy-Weinberg Equilibrium (HWE) p-value < 0.001
(4) minor allele frequency (MAF) < 0.01
(5) Mendelian error case > 3 families
(6) non-Mendelian error case > 3 families

Excluding those that did not meet the criteria, out of 871,166 SNPs probed from the 22 somatic chromosomes, 516,452 SNPs were retained for analysis. This data cleaning was done on

PedCheck[15] and Merlin software.

D. Statistical Analysis and Result Visualization

1. Descriptive analysis

To confirm that conventionally known associated risk factors play a role in our study population, Student's t-test, ANOVA, or chi-square test were applied as appropriate in SAS 9.3 for descriptive analyses.

2. Association Model and Test

For each SNP markers, one is to test whether the observed genotype and the phenotype are associated. So, consider the model below,

$$E(Y_{ij}) = \mu + \beta_g g_{ij} + \beta_x \mathbf{x}_{ij}$$

where μ is the population mean, β_g is the additive effect for each SNP, and β_x is a vector of covariate effects. Also, to allow for correlation between the observed phenotypes within each family, the variance-covariance matrix Ω_i for the i-th family is defined as below.

$$\Omega_{ijk} = \begin{cases} \sigma_a^2 + \sigma_g^2 + \sigma_e^2 & \text{if } j = k \\ \pi_{ijk}\sigma_a^2 + 2\varphi_{ijk}\sigma_g^2 & \text{if } j \neq k \end{cases}$$

Variance component parameters, σ^2_a, σ^2_g, and σ^2_e, account for linked major gene effects, polygenic effects, and environmental effects, respectively. Also, Π_{ijk} denotes identical-by-descent (IBD) sharing between individuals j and k at the location for the SNP being tested, and φ_{ijk} denotes the kinship coefficient between the same two individuals.

The association between individual SNPs and traits of interest (Hallux Valgus Angle) was analyzed using the FAmily-based Score Test for Association (FASTA) implemented in software Merlin with a command option of "--fastAssoc". FASTA used a linear mixed-model approximation to model the trait outcome, with genome data used to estimate kinship in order to account for relatedness including monozygotic twins. It also permitted adjustments to be made for age, sex, and BMI, using a normal distribution function.

--fastAssoc" option carries out a score test rather than a likelihood ratio test by "--assoc" option comparing two different models (Merlin, http://www.sph.umich.edu/csg/abecasis/Merlin/). When extensive computation procedures on a genome-wide scale are to be done, fitting a simple variance-components model works as an alternative to test the fitness. This model gives a vector of fitted values for each family, denoted as $E(\boldsymbol{y}_i)^{(\text{base})}$, and an estimate of the variance-covariance matrix for each family, $\Omega_i^{(\text{base})}$. Using these two, the score statistic can be defined as followed.

$$T^{\text{SCORE}} = \frac{\left\{\sum_i \left[\bar{g}_i - E(\bar{g}_i)\right]' \left[\Omega_i^{(\text{base})}\right]^{-1} \left[\mathbf{y}_i - E(y_i)^{(\text{base})}\right]\right\}^2}{\sum_i \left[\bar{g}_i - E(\bar{g}_i)\right]' \left[\Omega_i^{(\text{base})}\right]^{-1} \left[\bar{g}_i - E(\bar{g}_i)\right]}$$

Here, a vector for each individual's expected genotype scores in the i-th family is calculated conditionally on the given marker data, and denoted as $\bar{g}_i$. $E(\bar{g}_i)$ is a vector with same components that give the unconditional expectation of each genotype score, and this expectation equals 2p, which is twice the frequency of allele A at the SNP being tested. The value 2p arises from the Hardy-Weinberg equilibrium assumption in the population. So, for any i and j, we have

$$E(\bar{g}_{ij}) = E(g_{ij}) = 2\Pr(G_{ij} = \mathrm{A/A}) + \Pr(G_{ij} = \mathrm{A/a}) = 2p^2 + 2p(1-p) = 2p.$$

T^{score} is approximately distributed as χ^2 with 1 df since it only tests for significance of improvement in model fit by adding variables to the model, so it requires a single round of numerical optimization to estimate $\Omega_i^{(base)}$ and $E(y_i)^{(base)}$. And, it is important to note that the distribution of T^{score} will deviate from χ^2 when σ^2_a is large[16].

3. Result Visualization

Quantile-quantile (Q-Q) plots were generated in R version 3.0.3 (URL http://www.R-project.org/) to be on the lookout for any confounder in the GWA study. The QQ plot captures the expected distribution of association test statistics on the X-axis across the thousands of SNPs compared to the observed values on the Y-axis. If any deviation from the X=Y line is found, it represents a consistent and/or systematic difference between the expected and the observed. On the other had, a clean QQ plot should draw a solid line matching the X=Y slope till it sharply

curves towards the end, implying the small number of true associations among many thousands of unassociated SNPs.

Manhattan plots were generated using HaploView[17]. A Manhattan plot is a scattered plot, named after the view from the sky in Manhattan, NY with skyscrapers, to display data with a large number of data points, as in genome-wide association studies. Genomic coordinates are placed along the X-axis, and negative logarithms of the association p-value of each single nucleotide polymorphism are exhibited on the Y-axis.

Regional plots were createdusing LocusZoom (LocusZoom, URL https://statgen.sph.umich.edu/locuszoom/)in which negative logarithms of p-values were graphed against their chromosomal location by zooming in the region with an association peak among the whole SNP data for better observations on significant results. LocusZoom also visualizes linkage disequilibrium among SNPs in HapMap Phase II JPT + CHB population.

Linkage disequilibrium mapping was performed according to pairwise correlation structure analyzed by Haploview. The plot includes r2 values from the HapMap Phase II release 24 for JPT+CHB. LD plots display the probability that SNPs are in linkage disequilibrium. A higher probability (closer to 1) is colored darker, and white if it approaches to 0.5. LD plots were constructed through LocusZoom.

III. Results

A. Epidemiology of Hallux Valgus

The general characteristics of the 1265 study participants and prevalence of HV cases are described in Table 3: 38.7% (489 persons) were men, and the average age was 43.9 years for men and 44.1 for women. The prevalence of HV (HVA above 20°) was 16.4% (208 persons), and the crude (Cr) and standardized (St) prevalence of HV in twins were 13.9%(Cr) and 16.2% (St), and for non-twins 14.3%.

Table 3. General characteristics of study participants (n=1,265)

	HV angle mean(SD)	HV angle ≥ 20 (%)	Number of persons (%)
Total	14.36.5	208(16.4)	1265(100)
Height (cm)			
Men			
151-164	13.5(7.0)	14(14.6)	96(19.6)
165-169	13.3(6.1)	12(9.2)	130(26.6)
170-174	12.8(5.0)	9(6.5)	139(28.4)
174-191	12.5(5.4)	13(10.0)	124(25.4)
Women			
140-152	**16.6(7.7)**[a]	47(26.1)	180(23.2)
153-156	15.3(6.9)	47(26.3)	202(26.0)
157-160	**14.5(6.2)**[a]	31(15.8)	196(25.3)
161-175	**14.1(6.4)**[a]	35(17.7)	198(25.5)
Weight (kg)			
Men			
40-63	12.6(6.8)	10(10.0)	100(20.5)
64-70	12.8(5.5)	15(10.4)	144(29.5)
71-76	13.3(5.9)	12(10.4)	115(23.5)
77-135	13.5(5.1)	11(8.5)	130(26.6)
Women			
35-50	14.5(6.3)	26(15.5)	168(21.7)
51-57	15.6(7.3)	67(25.4)	264(34.0)
58-62	14.5(7.0)	21(18.9)	111(14.3)
63-97	15.2(6.4)	46(19.7)	233(30.0)

BMI (kg/m^2)			
≥25	**14.9(6.7)**[a]	76(18.5)	412(32.6)
<25	**14.0(6.4)**[a]	132(15.5)	853(67.4)
Sex[b]			
Men	**13.1±5.8**[b]	48(9.9)	489(38.7)
Women	**15.1±6.8**[b]	160(20.6)	776(61.3)
Age group[b]			
~29	**13.1±5.0**[b]	16(10.7)	150(11.9)
30-39	**13.5±5.6**[b]	48(12.2)	395(31.2)
40-49	**13.9±6.2**[b]	45(13.2)	316(25.0)
50-59	15.3±7.0	44(21.4)	206(16.3)
≥60	16.2±8.4	55(27.8)	198(15.7)
Twin	13.9±6.2	55(13.4)[16.2][1]	412(32.6)
Non-Twin	14.3±6.5	153(17.9)[18.3][1]	853(67.4)

[1]Figures in the square brackets indicate prevalence adjusted rates of HV standardized for age and sex. All other unspecified figures are crude mean and crude prevalence HV; [a] Group mean values of HV angle showing significant differences by ANOVA or t-test ($p < .05$); [b] Group mean values of HV angle showing marked statistically significant differences by ANOVA, chi-square test or t-test ($p < .001$); SD, standard deviation; H, hallux valgus (case or not); HVA, hallux valgus angle (as continuous measure); BMI, Quetelet's body mass index. Figures in bold indicate statistically significant findings.

B. GWAS results

A total of 516,452 genotyped SNPs were included in the analysis after quality control, and results shown by the QQ plot suggest the presence of multiple loci with significant effects (Figure 3). However, there are no significant SNPs associated with the trait upon the False Discovery Rate (FDR) control. SNPs with p-value less than 0.0001 are listed in Table 2 (n=77). The Manhattan plot is shown in Figure 4. No genotyped SNP met genome-wide significance criteria ($P < 5x10^{-8}$) for association with HVA increment.

Figure 5 presents the regional association plots for chromosome 1, 2, 7, and 16, whose segments contain those top SNPs. On each plot, the diamond indicates the top associated SNP on the chromosome, and colored dots the SNPs that show linkage disequilibrium (LD) within the range of r^2 from 0 to 1. LD plots for each of the four chromosomes are presented in Figure 6, each with colored arrow blocks pointing to the genotyped SNP.

Among the closest genes to the associated SNPs, shown in Table 4 and Figure 5, regions showing SNPs with p-value below $1x10^{-5}$ are near genes, DLX5, NTNG1, PRMT6, XYLT1, and CCDC85A that had been studied previously for their influences on limb formation, joint mobility, muscle function, and fibrogenesis, which may be a potential target not only for Hallux Valgus, but general foot development and disorders. In Table 5, previous studies on the listed genes are presented.

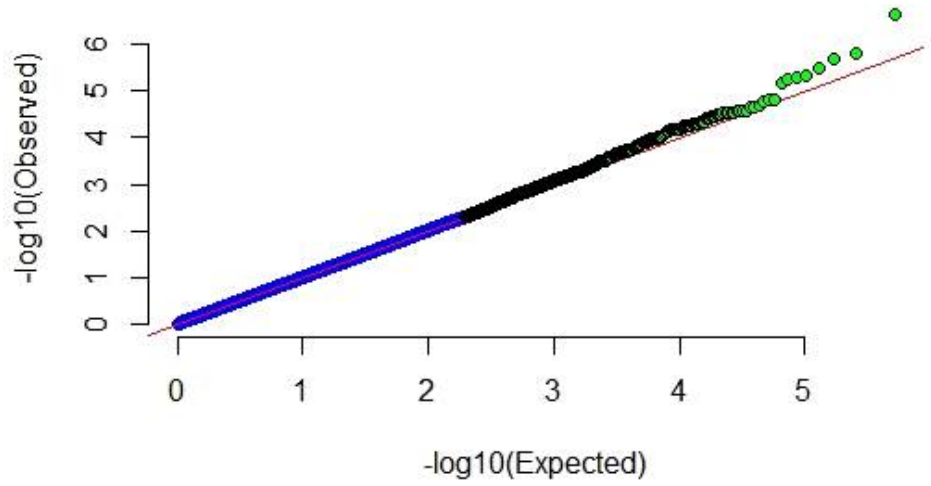

Figure 3. Quantile-Quantile plot

Table 4. 77 Significant SNPs associated with Hallux Valgus Angle deviation (p < 0.0001)

SNP	CHR	Cytoband	BP	MA	MAF	effect	SE	p-value	p-value after adjustment	closest gene
rs12669100	7	q21.3	96715879	T	0.389	-0.718	0.139	2.24E-07	0.115633603	ACN9,DLX5
rs1730852	1	p13.3	107655859	G	0.456	-0.678	0.141	1.52E-06	0.365475865	NTNG1,PRMT6
rs12443710	16	p12.3	18068778	T	0.411	0.671	0.141	2.12E-06	0.365475865	XYLT1
rs12614157	2	p16.1	57276155	T	0.16	-0.845	0.182	3.30E-06	0.399364954	CCDC85A
rs2894101	7	q21.3	96720040	G	0.356	-0.633	0.138	4.51E-06	0.399364954	ACN9,DLX5
rs12448862	16	p12.3	18068642	T	0.411	0.643	0.141	5.02E-06	0.399364954	XYLT1
rs10500394	16	p12.3	18069548	G	0.378	0.645	0.142	5.41E-06	0.399364954	XYLT1
rs10500395	16	p12.3	18069465	T	0.367	0.641	0.142	6.66E-06	0.43013996	XYLT1
rs1424649	2	p16.1	57370124	G	0.111	-0.881	0.204	1.50E-05	0.576893023	CCDC85A
rs12197033	6	q23.2	134854874	C	0.25	-0.698	0.161	1.53E-05	0.576893023	ALDH8A1
rs992611	4	q22.1	90899053	C	0.133	-0.903	0.21	1.75E-05	0.576893023	FAM190A
rs1881822	1	p13.3	107645245	T	0.489	-0.596	0.14	1.96E-05	0.576893023	NTNG1,PRMT6
rs7752190	6	q23.2	134858347	A	0.178	-0.688	0.162	2.12E-05	0.576893023	ALDH8A1
rs2717434	12	q15	71008547	A	0.433	0.607	0.143	2.16E-05	0.576893023	PTPRB
rs588441	5	q23.2	123766823	G	0.222	0.737	0.175	2.56E-05	0.576893023	ZNF608
rs6734341	2	p16.1	57390456	G	0.111	-0.827	0.197	2.60E-05	0.576893023	CCDC85A
rs1424647	2	p16.1	57369920	T	0.131	-0.845	0.201	2.69E-05	0.576893023	CCDC85A
rs952035	1	p13.3	107629329	C	0.489	-0.596	0.142	2.74E-05	0.576893023	NTNG1,PRMT6
rs11903107	2	p13.3	69039148	T	0.078	-1.035	0.247	2.82E-05	0.576893023	ARHGAP25
rs6714065	2	p13.3	69042461	T	0.078	-1.035	0.247	2.82E-05	0.576893023	ARHGAP25
rs9877722	3	q22.2	135635468	G	0	-2.283	0.546	2.92E-05	0.576893023	PPP2R3A

rs966115	2	p16.1	57387855	A	0.111	-0.832	0.199	2.97E-05	0.576893023	CCDC85A
rs13033180	2	p16.1	57119328	A	0.122	-0.779	0.187	3.04E-05	0.576893023	CCDC85A
rs1123146	22	q13.31	44929806	A	0.222	0.648	0.156	3.18E-05	0.576893023	KRT18
rs1762506	1	p13.3	107656356	C	0.489	-0.576	0.139	3.24E-05	0.576893023	NTNG1,PRMT6
rs2115603	2	p16.1	57270600	G	0.111	-0.862	0.207	3.26E-05	0.576893023	CCDC85A
rs7689553	4	q22.1	90910974	G	0.133	-0.855	0.206	3.43E-05	0.576893023	FAM190A
rs2115606	2	p16.1	57238715	A	0.111	-0.831	0.201	3.52E-05	0.576893023	CCDC85A
rs8061371	16	q23.1	77043677	A	0.033	-1.532	0.371	3.59E-05	0.576893023	MON1B
rs17189459	2	p16.1	57368572	G	0.135	-0.817	0.198	3.76E-05	0.576893023	CCDC85A
rs4807483	19	p13.3	3510005	A	0.122	0.909	0.221	3.88E-05	0.576893023	FZR1
rs6051342	20	p13	2715090	G	0.356	-0.629	0.153	4.01E-05	0.576893023	EBF4
rs1016884	20	q13.2	53485263	G	0.456	0.562	0.138	4.48E-05	0.576893023	RPL12
rs463129	5	q23.2	123765703	C	0.184	0.734	0.18	4.56E-05	0.576893023	ZNF608
rs6907469	6	q23.2	134865606	G	0.178	-0.668	0.164	4.69E-05	0.576893023	ALDH8A1
rs790663	10	q25.1	106787763	T	0.3	-0.639	0.157	4.85E-05	0.576893023	SORCS3
rs456409	5	q23.2	123765974	A	0.222	0.721	0.178	4.91E-05	0.576893023	ZNF608
rs11017438	10	q26.3	132547516	G	0.055	-1.274	0.314	4.93E-05	0.576893023	TCERG1L
rs7184877	16	q23.1	77031505	C	0.033	-1.437	0.355	5.16E-05	0.576893023	MON1B
rs829053	12	p12.1	25704772	C	0.078	1.106	0.273	5.22E-05	0.576893023	IFLTD1
rs6023672	20	q13.2	53497139	A	0.444	0.556	0.138	5.45E-05	0.576893023	RPL12
rs17189361	2	p16.1	57366620	G	0.122	-0.801	0.198	5.50E-05	0.576893023	CCDC85A
rs13097946	3	q22.3	135713622	A	0.011	-2.841	0.705	5.56E-05	0.576893023	PPP2R3A
rs13317459	3	q22.2	135670056	A	0	-2.839	0.705	5.62E-05	0.576893023	PPP2R3A

rs6774289	3	q22.3	135714105	A	0.011	-2.839	0.705	5.62E-05	0.576893023	PPP2R3A
rs6771997	3	q22.3	135731314	C	0.012	-2.838	0.705	5.62E-05	0.576893023	PPP2R3A
rs6064127	20	q13.2	53491566	T	0.489	0.562	0.14	5.72E-05	0.576893023	RPL12
rs2464046	1	p13.3	107665449	A	0.467	-0.565	0.141	5.96E-05	0.576893023	NTNG1,PRMT6
rs17058743	3	p14.3	58009372	A	0.478	-0.57	0.142	5.97E-05	0.576893023	FLNB
rs7513659	1	q42.13	230226386	T	0.044	-1.906	0.477	6.33E-05	0.576893023	GALNT2
rs13285280	9	q21.13	79163629	T	0.178	-1.001	0.25	6.37E-05	0.576893023	PRUNE2
rs6127303	20	q13.2	53481127	T	0.471	0.55	0.138	6.39E-05	0.576893023	RPL12
rs11043113	11	p15.3	11167748	A	0.033	1.832	0.459	6.43E-05	0.576893023	GALNTL4
rs1526084	7	q21.3	96697572	T	0.333	-0.573	0.144	6.59E-05	0.576893023	ACN9,DLX5
rs16889787	8	p11.21	40496978	T	0.389	0.592	0.148	6.63E-05	0.576893023	ZMAT4
rs2722419	8	p11.21	40490493	G	0.389	0.592	0.148	6.63E-05	0.576893023	ZMAT4
rs2281747	9	p24.1	8465598	C	0.267	0.621	0.156	6.65E-05	0.576893023	PTPRD
rs2619480	12	p12.1	25668635	G	0.122	0.942	0.236	6.69E-05	0.576893023	IFLTD1
rs13113035	4	q22.1	90958920	G	0.122	-0.831	0.209	6.81E-05	0.576893023	FAM190A
rs2465810	12	q15	71015280	G	0.389	-0.601	0.151	6.82E-05	0.576893023	PTPRB
rs624699	6	p12.3	51514608	A	0.144	0.78	0.196	6.84E-05	0.576893023	PKHD1
rs6696357	1	q23.1	158369319	A	0.266	-0.644	0.162	7.11E-05	0.576893023	OR10K2
rs6672459	1	q23.1	158371678	C	0.344	-0.644	0.162	7.12E-05	0.576893023	OR10K2
rs10122018	9	p21.3	24677263	T	0.022	-1.821	0.459	7.15E-05	0.576893023	TUSC1
rs4538607	5	q23.2	123790231	T	0.144	0.88	0.222	7.55E-05	0.600117224	ZNF608
rs12815089	12	p12.1	22954932	C	0.189	-0.679	0.173	8.27E-05	0.634172676	SOX5
rs12444719	16	q21	61070859	G	0.411	0.599	0.152	8.32E-05	0.634172676	CDH8

rs17119214	10	q25.1	107371475	G	0	-2.816	0.716	8.35E-05	0.634172676	YWHAZ
rs585273	1	q43	236636159	A	0.444	0.54	0.138	8.61E-05	0.635531075	EDARADD
rs6069021	20	q13.2	53493371	G	0.367	0.56	0.143	8.61E-05	0.635531075	RPL12
rs9828717	3	p14.3	57998986	G	0.465	-0.551	0.141	9.22E-05	0.642630614	FLNB
rs228274	17	q12	36909061	T	0.122	0.829	0.212	9.43E-05	0.642630614	PSMB3
rs2599657	8	p11.21	40500792	T	0.389	0.588	0.151	9.74E-05	0.642630614	ZMAT4
rs10231759	7	q36.1	150512172	A	0.3	-0.65	0.167	9.75E-05	0.642630614	ABP1
rs1451680	8	p11.21	40491838	A	0.389	0.578	0.148	9.80E-05	0.642630614	ZMAT4
rs7296326	12	q14.1	62667727	C	0.022	-1.767	0.454	9.90E-05	0.642630614	USP15
rs3857964	8	p11.21	40836998	A	0.233	-0.601	0.154	9.96E-05	0.642630614	SFRP1

*Note: BP, base pair; MA, minor allele; MAF, minor allele frequency; SE, standard error

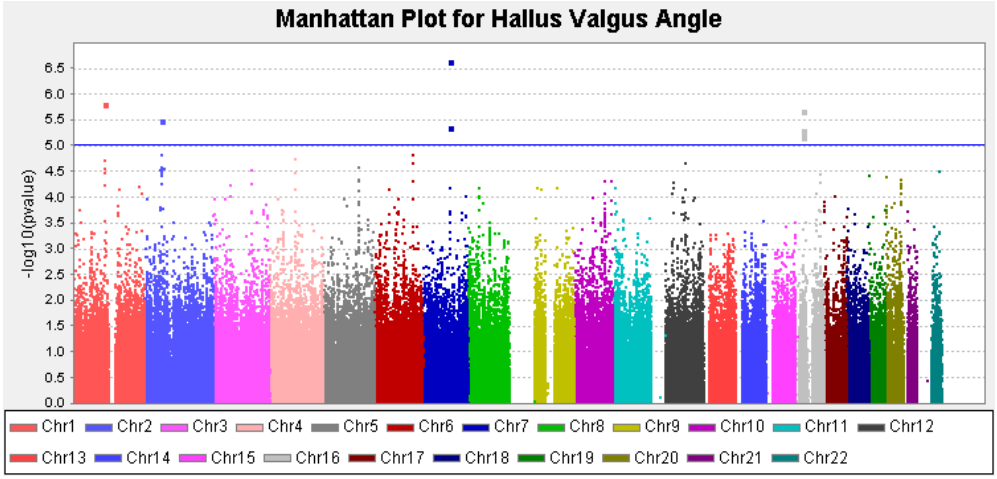

Figure 4. Manhattan plot for Hallux Valgus Angle deviation. the horizontal line corresponds to a p-value of 5.0×10^{-5}.

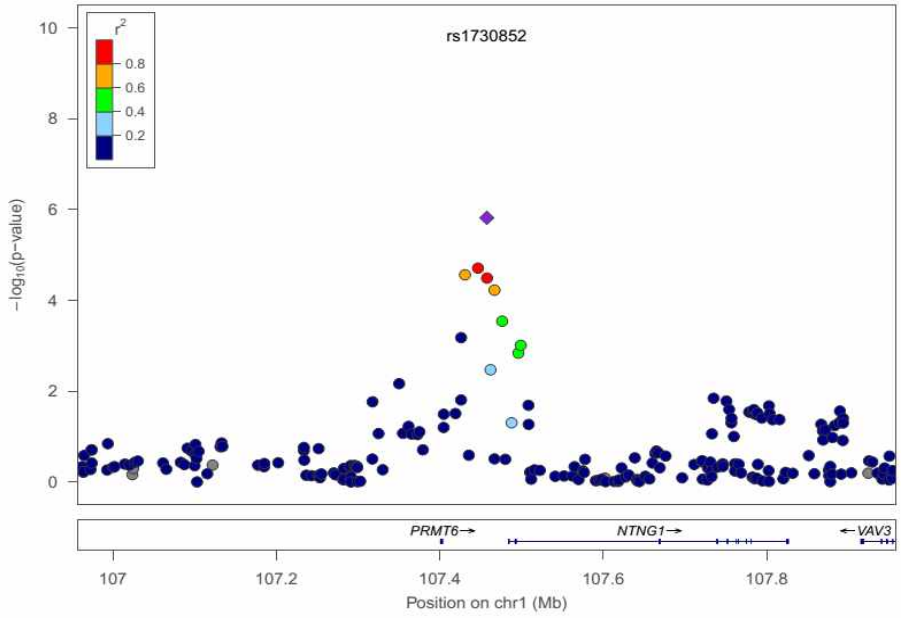

Figure 5-1. Regional plot for chromosome 1 with rs1730852 (p=1.52E-06)

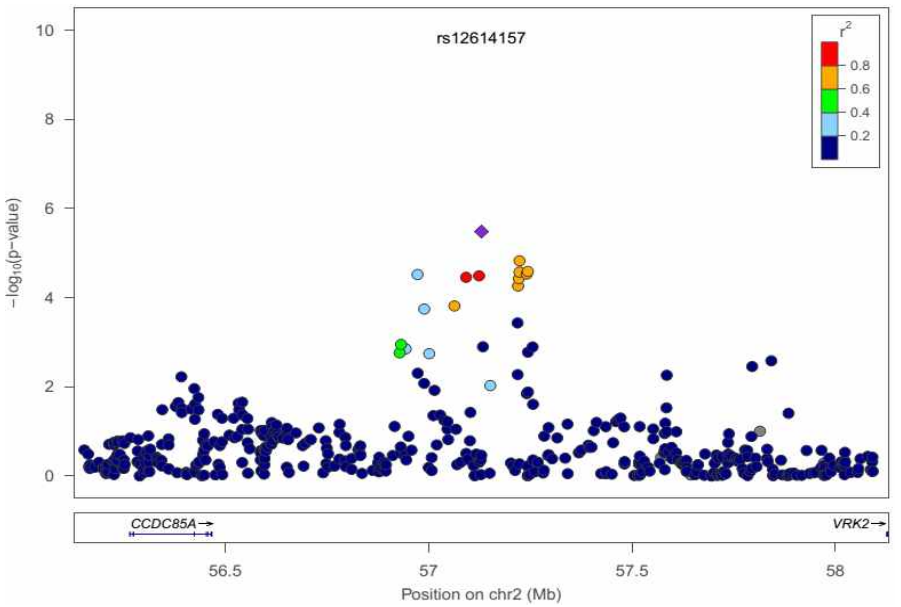

Figure 5-2. Regional plot for chromosome 2 with rs12614157 (p=3.30E-06)

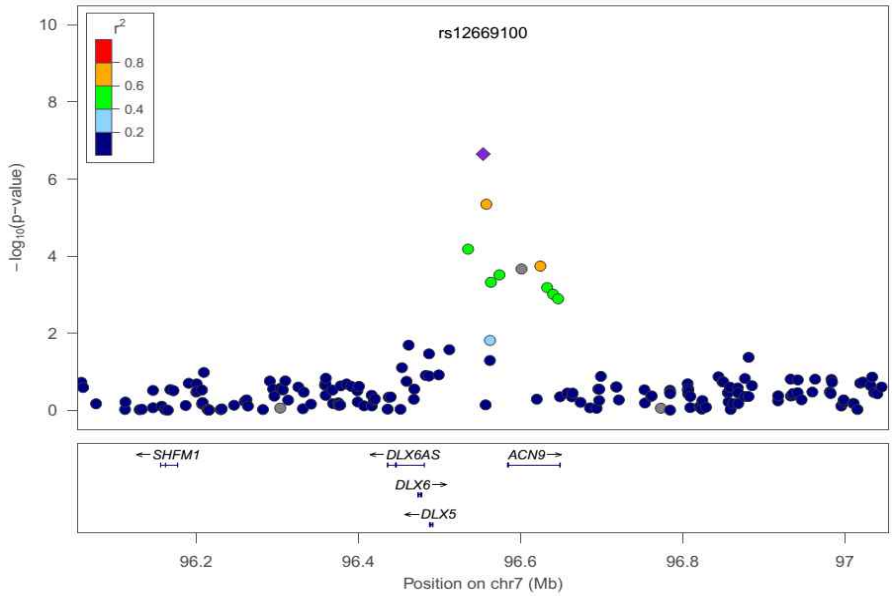

Figure 5-3. Regional plot for chromosome 7 with rs12669100 (p=2.24E-07)

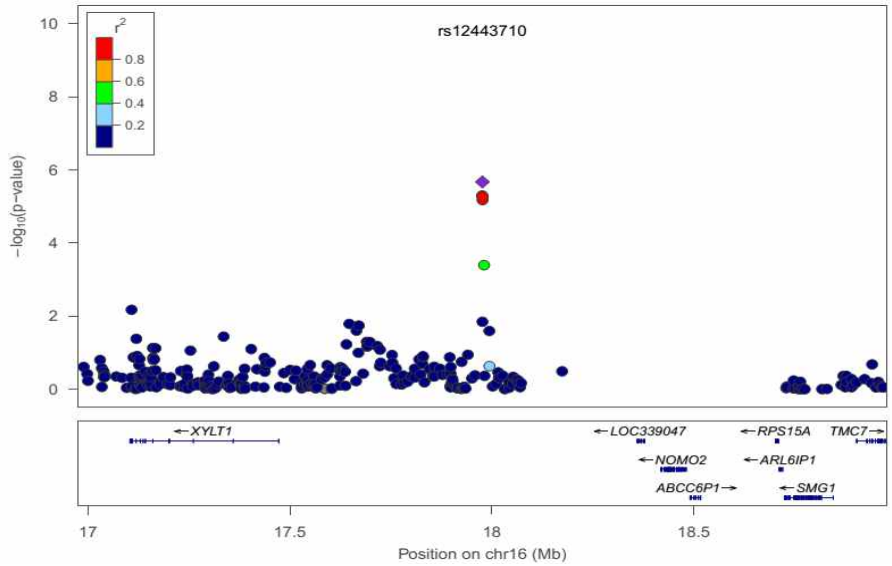

Figure 5-4. Regional plot for chromosome 16 with rs12443710 (p=2.12E-06)

Figure 5. Regional plots for chromosome 1, 2, 7, and 16.

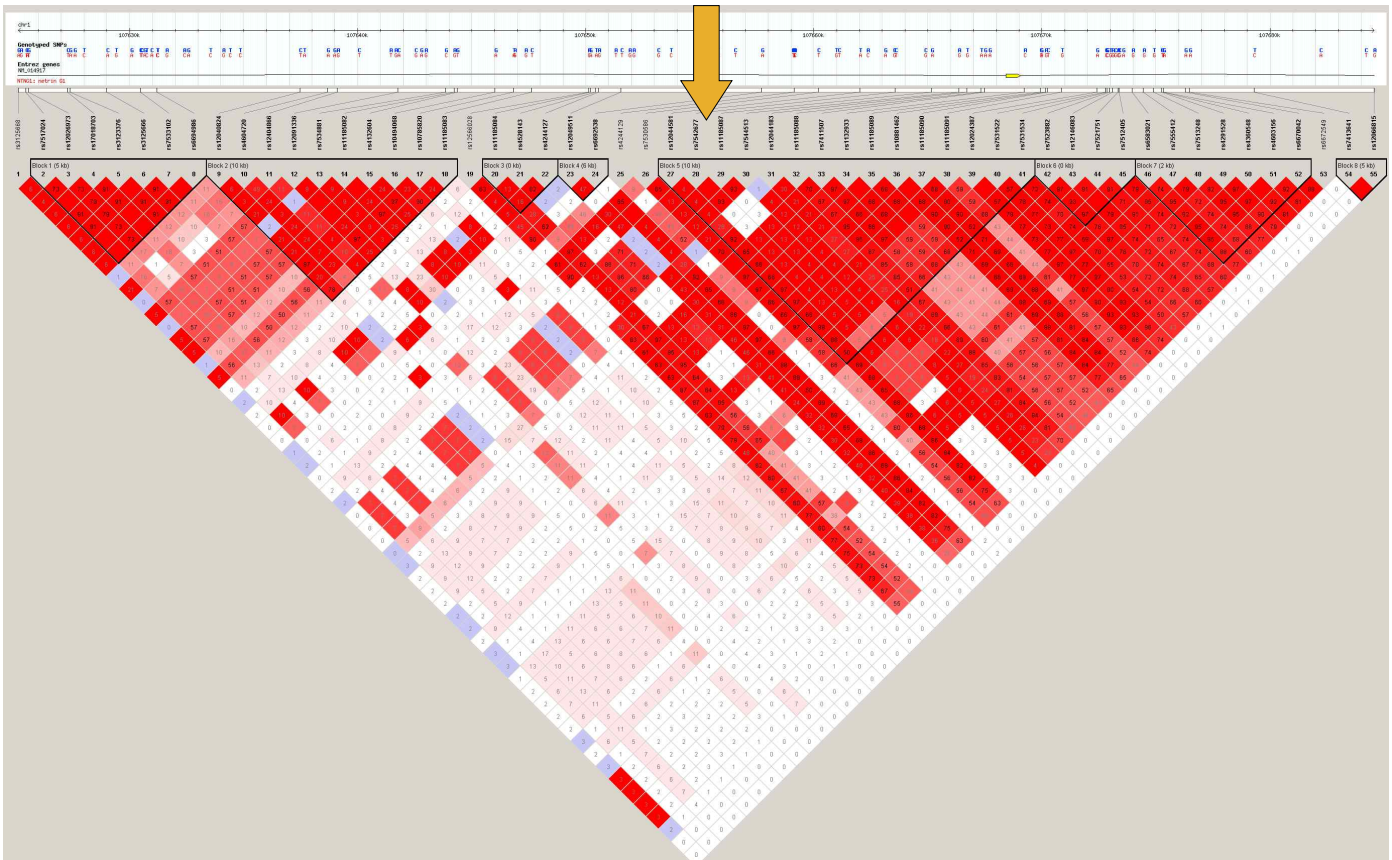

Figure 6-1. LD plot for Chromosome 1 with regions including rs12669100 (p=2.24e-07)

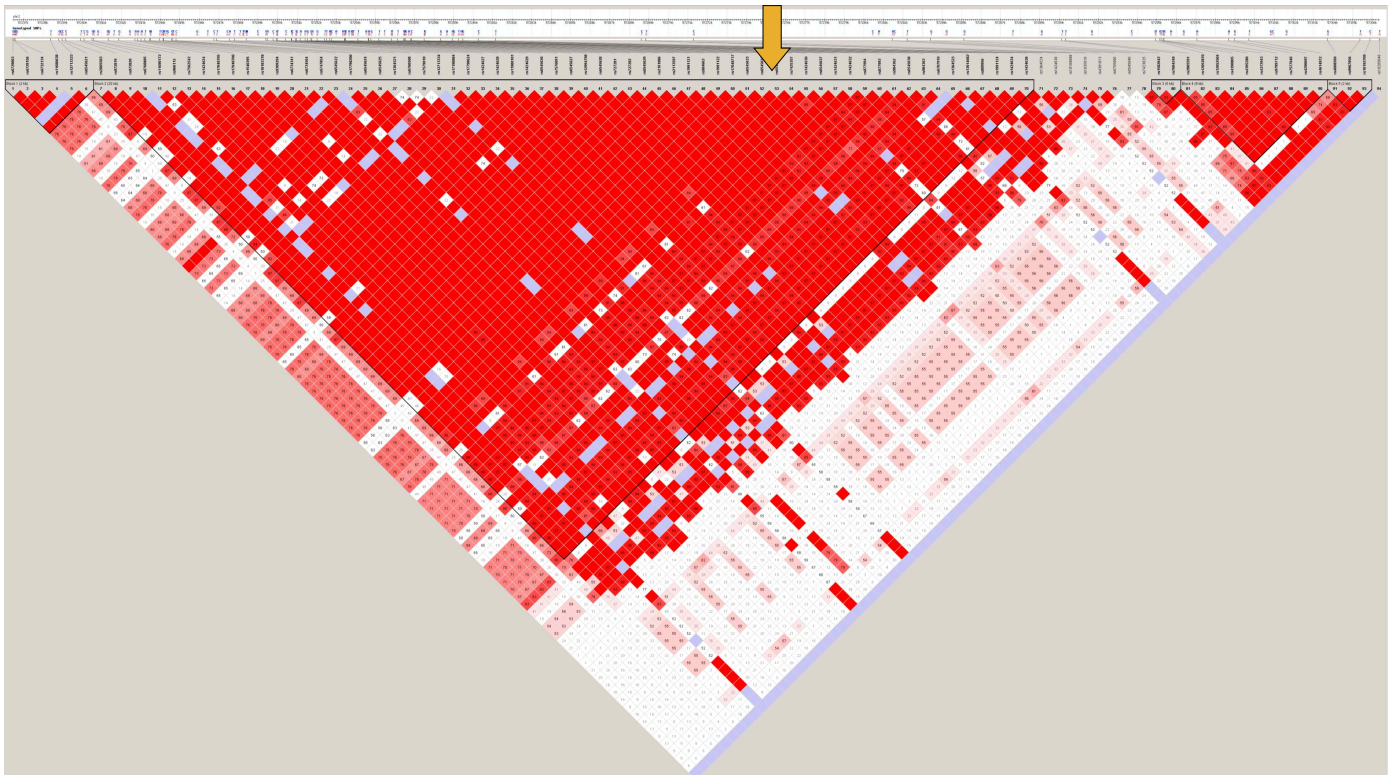

Figure 6-2. LD plot for Chromosome 2 with regions including rs12614157 (p=3.30e-06)

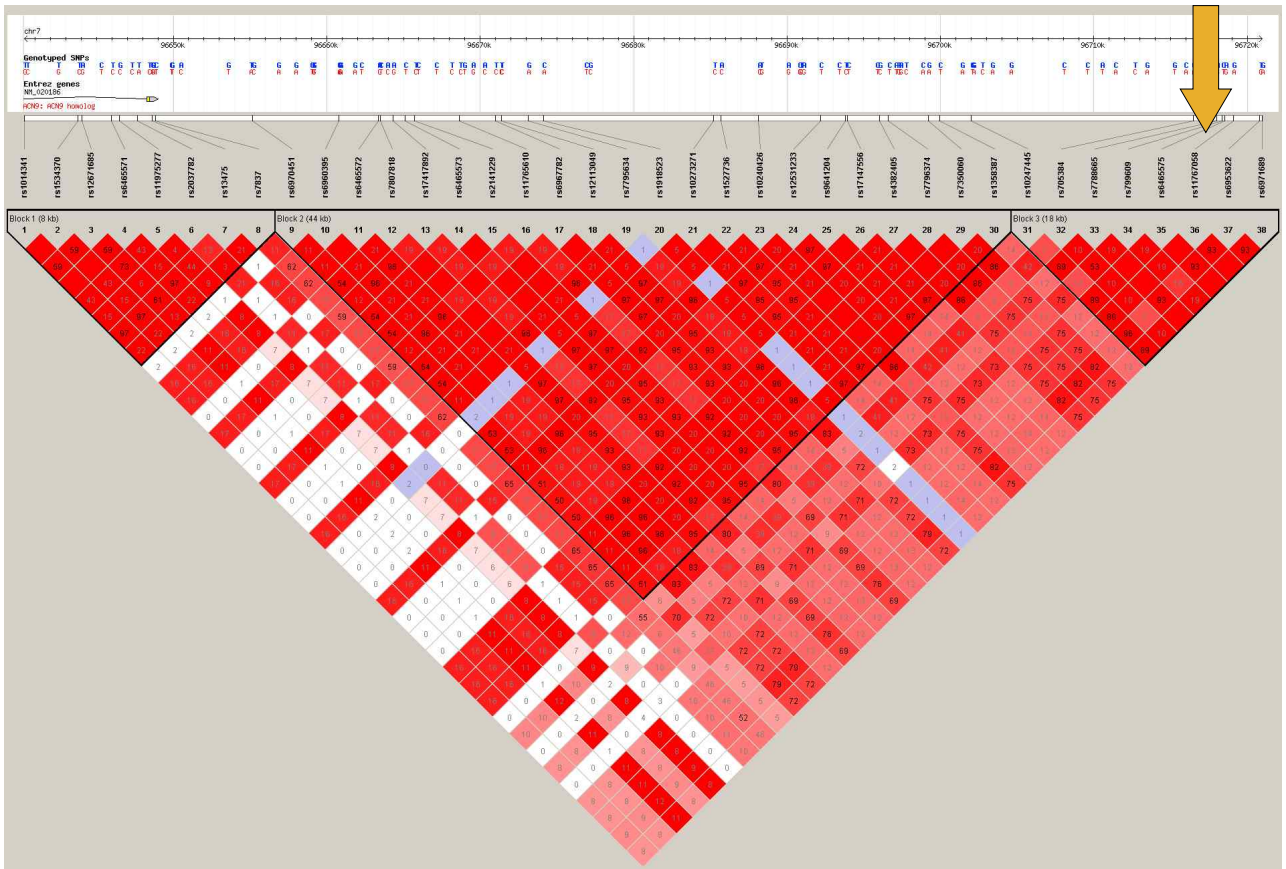

Figure 6-3. LD plot for Chromosome 7 with regions including rs12669100 (p=2.24e-07)

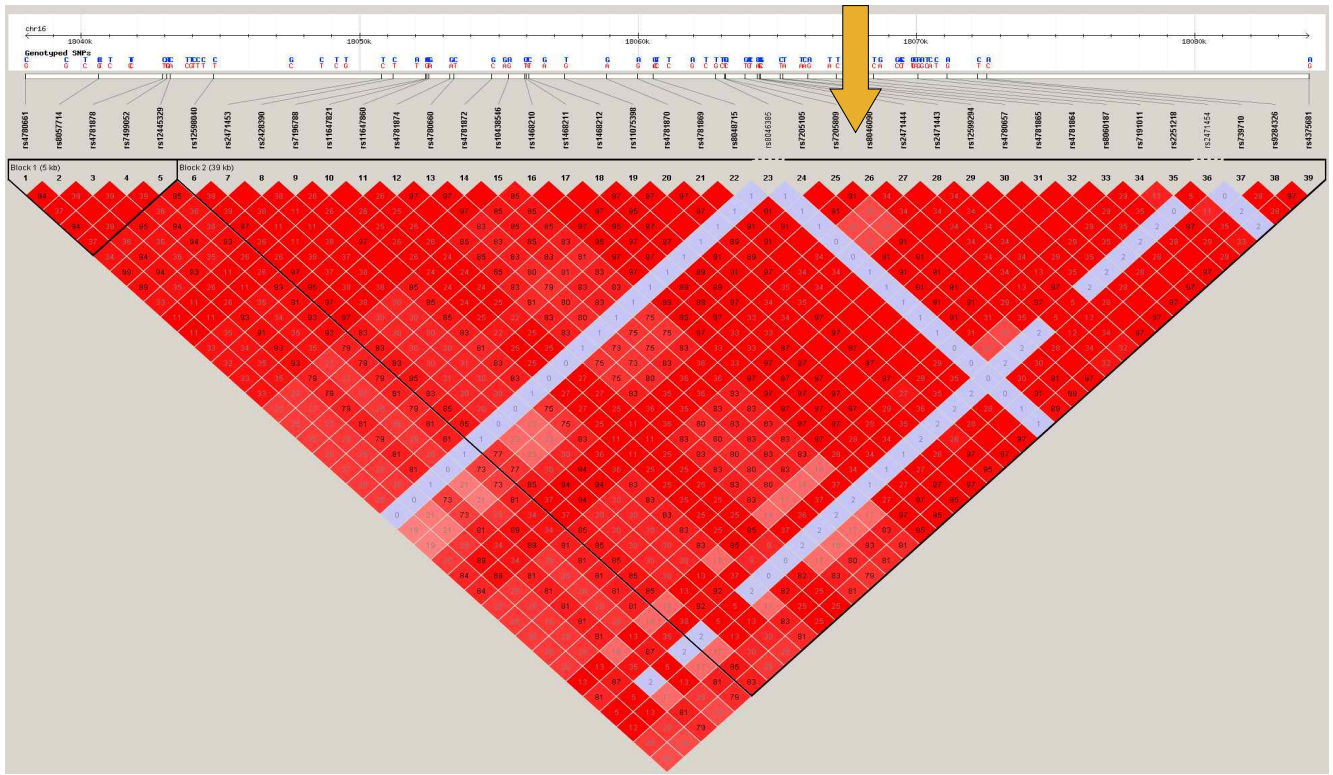

Figure 6-4. LD plot for Chromosome 16 with regions including rs12443710 (p=2.12e-06)

Figure 6. Numbers within diamonds indicate the r^2 values. For those SNPs without any listed value, r^2=1. The arrows point to the genotyped SNP.

Table 5. Name and associated disease/functions of each of the 37 genes from Table 4.

Gene	Name	Associated diseases	Function	Reference
ABP1	Amine Oxidase, Copper Containing 1		Catalyzes degradation of histamine, spermine, and spermidine, substances involved in allergic and immune responses, cell proliferation, tissue differentiation, tumor formation, and possibly apoptosis.	GeneCards
ACN9	Acetate Nonutilizing 9	acne, alcohol dependence	Encodes a protein of the mitochondrial intermembrane space that is required for growth on acetate and other nonfermentable carbon sources. Mutations in ACN9 are in charge of increased leves of enzymes for the glyoxylate cycle, gluconeogenesis, and acetyl-CoA metaboism.	18, 19
ALDH8A1	Aldehyde Dehydrogenase 8 Family, Member A1	Sjogren-Larsson Syndrome, Beta thalassemia	Converts 9-cis-retinal to 9-cis-retinoic acid	RefSeq, Jul 2010
ARHGAP25	Rho GTPase Activating Protein 25	spasticity	Encodes negative regulators of Rho GTPases, which are implicated in actin remodeling, cell polarity, and cell migration	20
CCDC85A	Coiled-Coil Domain Containing	Neurofibromatosis	Encodes coiled coil stretch protein	GeneCards
CDH8	Cadherin 8, Type 2	Cystic fibrosis, Parkinson's disease	Encodes a type 2 cadherin, which is a calcium-dependent cell adhesion protein connecting cells in a homophilic manner	UniProtKB
DLX5	Distal-Less Homeobox 5	Split hand / foot, ectrodactyly, cleft hand, sensorineural hearing loss	Encodes protein that plays a role in bone development and fracture healing	EntrezGene, GeneCards

EBF4	Early B-Cell Factor 4	Gestational choriocarcinoma	Plays a role in neural development and B-cell maturation	21
EDARADD	EDAR-Associated Death Domain	Ectodermal dysplasia	Encodes a protein required for the development of hair, teeth and other ectodermal derivatives	RefSeq, Jul 2008
FAM190A	family with sequence similarity 190, member A	choromosomal instability in cancer	Its internal rearrangements result in disruption of tumor-suppressor genes	22
FLNB	Filamin B, Beta	Larsen Syndrome	Encodes a member of the filamin family that interacts with glycoprotein lb alpha as part of the process to repair vascular injuries. Mutaions are found in Larsen syndrome cuasing abnormal development of the bones so as to lead to clubfoot and numerous joint dislocations affecting the hips, knees and elbows; flexible joints; and a distinctive appearance of the face, hands and feet.	RefSeq, Nov 2009; GeneCards
FZR1	Fizzy/Cell Division Cycle 20 Related 1	Cleidocranial dysplasia	Key regulator of ligase activity of the anaphase promoting complex/cyclosome (APC/C) that controls entry into mitosis in cell cycles	
GALNT2	UDP-N-Acetyl-Alpha-D-Galactosamine: Polypeptide N-Acetylgalactosaminyltransferase 2	hypertriglyceridemia, tonsilitis	Catalyzes the initial reaction in O-linked oligosaccharide biosynthesis, the transfer of an N-acetyl-D-galactosamine residue to a serine or threonine residue on the protein receptor.	UniProtKB; GeneCards
GALNTL4	UDP-N-acetyl-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-like 4	Celiac disease	protein amino acid glycosylation	AceView

IFLTD1	Intermediate Filament Tail Domain Containing 1	lung adnoma	structural molecule activity; localizes in nucleus	GeneCards; AceView
KRT18	Keratin 18	Gallbladder sarcoma, Schneiderian carcinoma	scaffold protein binding; structural molecule activity	GeneCards
MON1B	MON1 Secretory Trafficking Family Member B	Detrusor sphincter dyssynergia, herpes simplex	protein binding	GeneCards
NTNG1	Netrin G1	Rett Syndrome	Acts as axon guidance cues during vertebrate nervous system development	23
OR10K2	Olfactory Receptor, Family 10, Subfamily K, Member 2	Anterograde amnesia, peroneal neuropathy	Interacts with odorant molecules in the nose	
PKHD1	Polycystic Kidney And Hepatic Disease 1	congenital hepatic fibrosis, polycystic kidney and hepatic disease	Encodes the protein to have multiple copies of an immunoglobulin-like plexin-transcription-factor domain.	RefSeq, Jul 2008
PPP2R3A	Protein Phosphatase 2, Regulatory Subunit B'', Alpha	retinoblastoma, breast cancer	encodes one of the regulatory subunits of the protein phosphatase 2	RefSeq, Jun 2010
PRMT6	Protein Arginine Methyltransferase 6	glycogen storage disease	catalyze the sequential transfer of methyl group from S-adenosyl-L-methionine to the side chain nitrogens of arginine residues within proteins, to form methylated arginine derivatives and S-adenosyl-L-homocysteine.	RefSeq, Sep 2011
PRUNE2	Prune Homolog 2 (Drosophila)	Leiomyosarcoma, gastrointestinal stromal tumor	pyrophsphatase activity, metal ion binding	GeneCards

PSMB3	Proteasome (Prosome, Macropain) Subunit, β 3	Huntington's disease, Parkinson's disease	threonine-type endopeptidase activity	RefSeq, Sep 2013
PTPRB	Protein Tyrosine Phosphatase, Receptor Type, B	Hyperkalemic periodic paralysis, gastric ulcer	protein binding and transmembrane receptor protein tyrosine phosphatase activity	
PTPRD	Protein Tyrosine Phosphatase, Receptor Type, D	melanoma astrocytoma syndrome, plague	transmembrane receptor protein tyrosine phosphatase activity and receptor binding	RefSeq, Jan 2010
RPL12	Ribosomal Protein L12	Cleft lip, Intrahapatic cholangiocarcinoma	structural constituent of ribosome and RNA binding	RefSeq, Jul 2008
SFRP1	Secreted Frizzled-Related Protein 1	familial chronic lymphocytic leukemia, clear cell renal cell carcinoma	Determines the polarity of photoreceptor cells in the retina	RefSeq, Sep 2009
SORCS3	Sortilin-Related VPS10 Domain Containing Receptor 3	Alzheimer's disease	neuropeptide receptor activity; expressed in the central nervous system	RefSeq, Jul 2008
SOX5	SRY (Sex Determining Region Y)-Box 5	developmental and speech delay due to sox5 deficiency, 12p12.1 microdeletion syndrome	transcription regulatory region DNA binding; sequence-specific DNA binding transcription factor activity	RefSeq, Jul 2008
TCERG1L	Transcription Elongation Regulator 1-Like	insulin resistance, Crohn's disease	protein binding	AceView
TUSC1	Tumor Suppressor Candidate 1	non-small cell lung carcinoma, lung cancer	Downregulates in non-small-cell lung cancer and small-cell lung cancer cell lines, suggesting that it may play a role in lung tumorigenesis	RefSeq, Jul 2008

USP15	Ubiquitin Specific Peptidase 15	spinocerebellar ataxia type 3, spinocerebellar ataxia	transforming growth factor beta signalling through deubiquitination of receptor-activated SMAD transcription factors.	RefSeq, Nov 2011
XYLT1	Xylosyltransferase 1	Pseudoxanthoma Elasticum, Modifier of Severity of	Catalyzes the first step in biosynthesis of glycosaminoglycan	UniProtKB; RefSeq, Nov 2009
YWHAZ	Tyrosine 3-Monooxygenase/ Tryptophan 5-Monooxygenase Activation Protein, Zeta Polypeptide1	paranoid schizophrenia, and balanitis	Regulates insulin sensitivity	GeneCards; REfSeq, Oct 2008
ZMAT4	Zinc Finger, Matrin-Type 4		DNA binding; zinc ion binding	GeneCards
ZNF608	Zinc Finger Protein 608	renal cell carcinoma, prostate cancer	zinc ion binding	GeneCards

IV. Discussion

In this study, a genome-wide association study for the discovery of loci associated with hallux valgus angle deviation was conducted through Merlin, software performing the family-based association score test, recognizing monozygotic twins. Hallux valgus is known to have older age, female sex, and higher BMI as risk factors, and in this study population, too, they were significantly associated with the metatarsal deviation. Although many variants were to be anticipated to be associated with the trait of interest looking from the quantile-quantile plot, the top SNP (rs12669100, $p=2.24x10^{-7}$) did not reach its genome-wide significance. Among the associated SNPs identified, the regions of the leading SNPs comprise DLX5, NTNG1, PRMT6, XYLT1, and CCDC85A genes: DLX5 is known to be associated with split-hand/split-foot malformation whose symptoms are syndactyly and aplasia/hypoplasia of the phalanges and metatarsals; NTNG1 with Rett Syndrome, the symptoms of which includes restricted joint mobility, joint stiffness, and clinodactyly of fifth finger; PRMT6 with glycogen storage disease that affects glycogenosis and breakdown within muscles; XYLT1 with Pseudoxanthoma Elasticum accompanied by accumulation of calcium and other minerals in the elastic fibers of the body; and CCDC85A with neurofibromatosis affiliated deformed bones. These imply that the identified variants may be a target to hallux valgus treatment and general foot development.

With some possible implications, there are several

limitations of this study to be acknowledged. Since this study was not primarily targeted to hallux valgus as a dichotomous trait, it does not necessarily imply associated SNPs with the deformation itself. And, no single SNP satisfied FDR-adjusted statistical significance. The sample size for the genome-wide analysis may have limited statistical power. Also, the findings may not be generalizable to other ethnic groups, since the prevalence of the deformity is thought to be different across the populations.

In addition, if there were information on age at onset of constricting footwear use, analysis on genotype-by-environment interactions (GxE) could be possible. However, the survey reports only had the current habit of wearing shoes, and interestingly it shows a reverse direction with significance - constricting shoe wear is associated with decreasing hallux valgus angle, rather than the anticipated, of which its reason can be guessed as those who currently have angle deviation may be likely to avoid wearing high-heeled or constricting shoes.

In conclusion, this study observed statistically suggestive genetic determinants associated with hallux valgus angle deviation, which can further our understanding of biologic mechanisms underlying hallux valgus and its etiology toward early prevention.

V. References

1. McGeoch, P.D., *Does cortical reorganisation explain the enduring popularity of foot-binding in medieval China?* Med Hypotheses, 2007. 69(4): p. 938-41.
2. Pan, Y., et al., *A study on bone mass in elderly chinese foot-binding women.* Int J Endocrinol, 2013. 2013: p. 351670.
3. Lee, C.H., et al., *Genetic influences on hallux valgus in Koreans: the healthy twin study.* Twin Res Hum Genet, 2014. 17(2): p. 121-6.
4. Nix, S., M. Smith, and B. Vicenzino, *Prevalence of hallux valgus in the general population: a systematic review and meta-analysis.* J Foot Ankle Res, 2010. 3(1): p. 21.
5. Hill, C.L., et al., *Prevalence and correlates of foot pain in a population-based study: the North West Adelaide health study.* J Foot Ankle Res, 2008. 1(2): p. 1-7.
6. Menz, H.B., E.L. Barr, and W.J. Brown, *Predictors and persistence of foot problems in women aged 70 years and over: A prospective study.* Maturitas, 2011. 68(1): p. 83-87.
7. Nishimura, A., et al., *Prevalence of hallux valgus and risk factors among Japanese community dwellers.* Journal of Orthopaedic Science, 2013: p. 1-6.
8. Hannan, M.T., et al., *High heritability of hallux valgus and lesser toe deformities in adult men and women.* Arthritis care & research, 2013. 65(9): p. 1515-1521.
9. Hardy, R. and J. Clapham, *Observations on hallux valgus.* Journal of Bone & Joint Surgery, British Volume, 1951. 33(3): p. 376-391.
10. MITCHELL, C.L., et al., *Osteotomy-bunionectomy for hallux valgus.* The Journal of Bone & Joint Surgery, 1958. 40(1): p. 41-60.
11. Piqu□Vidal, C., M.T. Sol□ and J. Antich, *Hallux valgus inheritance: pedigree research in 350 patients with bunion deformity.* The Journal of foot and ankle surgery, 2007. 46(3): p. 149-154.
12. Kilmartin, T., R. Barrington, and W. Wallace, *Metatarsus Primus Varus.* Journal of Pediatric Orthopaedics, 1992. 12(3): p. 415.
13. Coughlin, M.J., *Juvenile hallux valgus: etiology and treatment.* Foot & Ankle International, 1995. 16(11): p. 682-697.
14. Sung, J., et al., *Healthy Twin: a twin-family study of Korea--protocols and current status.* Twin Res Hum Genet, 2006. 9(6): p. 844-8.
15. O'Connell, J.R. and D.E. Weeks, *PedCheck: a program for identification of genotype incompatibilities in linkage analysis.* Am J Hum Genet, 1998. 63(1): p. 259-66.
16. Chen, W.M. and G.R. Abecasis, *Family-based association tests for genomewide association scans.* Am J Hum Genet, 2007. 81(5): p. 913-26.
17. Barrett, J.C., *Haploview: Visualization and analysis of SNP genotype data.* Cold Spring Harb Protoc, 2009. 2009(10): p. pdb.ip71.
18. McCammon, M.T., *Mutants of Saccharomyces cerevisiae with defects in acetate metabolism: isolation and characterization of Acn- mutants.* Genetics, 1996. 144(1): p. 57-69.

19. Dennis, R.A. and M.T. McCammon, *Acn9 is a novel protein of gluconeogenesis that is located in the mitochondrial intermembrane space.* Eur J Biochem, 1999. 261(1): p. 236-43.
20. Katoh, M. and M. Katoh, *Identification and characterization of ARHGAP24 and ARHGAP25 genes in silico.* Int J Mol Med, 2004. 14(2): p. 333-8.
21. Wang, S.S., A.G. Betz, and R.R. Reed, *Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-based direct selection.* Mol Cell Neurosci, 2002. 20(3): p. 404-14.
22. Scrimieri, F., et al., *FAM190A rearrangements provide a multitude of individualized tumor signatures and neo-antigens in cancer.* Oncotarget, 2011. 2(1-2): p. 69-75.
23. Nakashiba, T., et al., *Netrin-G1: a novel glycosyl phosphatidylinositol-linked mammalian netrin that is functionally divergent from classical netrins.* J Neurosci, 2000. 20(17): p. 6540-50.

국문초록

무지외반증과 관련된 유전요인 및 관련요인 분석

서울대학교 보건대학원
보건학과 유전체역학전공
이 수 지

무지외반증은 엄지발가락의 뿌리 부분인 제1중족 발가락 관절을 기준으로 족지 쪽의 뼈가 안으로 회내하는 삼차원적 변형으로서 나이, 성별, 체질량지수, 하이힐 등의 신발 코가 좁고 굽이 높은 신발을 자주 이용하는 경우와 같은 후천적 요인과 함께 유전적 요인의 영향을 받는다고 알려져 있다. 선행 연구들에 따르면 무지외반증의 유전율은 0.51 (한국인), 0.29-0.89 (백색인종)으로 높게 보고되었다. 본 연구의 목적은 무지외반각의 증가에 기인하는 관련 유전 변이를 탐색하는 데에 있다.

본 연구는 한국인 가족-쌍둥이 코호트 연구 참여자들 중 족부건강 설문에 응답하고, 족부 X-ray 촬영에 응한 1265명을 대상으로 한다. 관련 유전자 발견을 위해서 전장유전체 연관성 분석을 시행하였다. 설문 응답자 중, 995명의 genotype 분석은 Affymetrix Genome-Wide Human SNP Array 6.0을 기반으로 수행되었다. 연구대상인구집단에서 확인된 연령, 성별, 체질량지수와 같은 위험 요인을 보정하여 전장 유전체 연관성 분석을 실시하였다.

연구 결과, 본 연구에서 무지외반각의 증가와 가장 높은 유의성을 보인 단일 염기 다형성 (SNP: single nucleotide polymorphism)은 rs12669100였으며, 이는 다중비교의 문제를 해결하기 위한 FDR 보정 후에는 0.11의 유의확률에 그쳤다. 통계적으로 유의미한 유전변이들이 포함되거나 가까이 위치한 유전자는 DLX5, NTNG1, PRMT6, XYLT1, CCDC85A 이었으며, 각각 수, 족지 형성, 근육-신경 형성 등에 영향을 미치는 것으로 알려져 있다.

본 연구는 한국인 집단에서 무지외반각의 편향에 대한 연관 유전변이를 발견하였고, 근래의 외모지상주의에 따른 여성의 발 건강에 대한 높아지는 사회적 관심에 따라 보건학적인 관점에서 의의를 가진다. 향후 이 연구에서 밝혀진 유전변이들이 복제연구를 통해 재차 확인될 것으로 기대하며, 위험군에 대해서 신발착용행태와 유전적 원인 간의 상호 작용에 대한 분석도 필요하겠다.

주요어: 무지외반증, 전장 유전체 연관성 분석, 단일염기 다형성, 유전변이, 가족 기반 연관성 분석

학 번: 2012-21895